Renewable and Sustainable Energy Reviews 27 (2013) 613–621

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine

Jie Yan*, Yongqian Liu, Shuang Han, Meng Qiu

State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of Renewable Energy, North China Electric Power University, Beijing 102206, China

ARTICLE INFO

Article history:
Received 8 September 2012
Received in revised form
12 July 2013
Accepted 14 July 2013
Available online 7 August 2013

Keywords:
Wind power
Grouping forecasts
Uncertainty analysis
Optimized relevance vector machine

ABSTRACT

Relevance vector machine, a sparse probabilistic learning machine based on the kernel function, has excellent ability of prediction and generalization. It is proposed by this paper that the optimized relevance vector machine (ORVM) is a wind power interval forecasting model which is able to provide a certain prediction value and its possible fluctuation range at a given confidence level. The proposed model characterizes in insufficient sample training and uncertainty analysis and is greatly suitable to most of wind farms in China (newly built or large scale wind farms). First, a grouping mechanism has been used to divide wind turbines into several groups to establish forecasting model separately. Second, a selection method properly taking the characteristics of NWP error distribution into consideration was presented to improve forecasting accuracy of each group. Third, the parameters of the kernel function and initial value of iteration are determined by particle swarm optimization to further enhance forecasting accuracy. Two wind farms in China are involved in the process of primary data collection. The performance data obtained from ORVM models are tested against the predicted data generated by GA–ANN and SVM. Results show that the proposed model has better prediction accuracy, wider application scope and more efficient calculation.

Contents

1. Introduction . . . 613
2. Theory of relevance vector machine [14] . . . 614
3. RVM-based wind power interval forecasting model . . . 615
   3.1. Model structure . . . 615
   3.2. Grouping of wind turbines . . . 615
   3.3. Selection of training samples . . . 616
   3.4. Optimization of kernel function . . . 617
   3.5. Steps of grouping forecasts . . . 617
4. Case study . . . 618
   4.1. Data . . . 618
   4.2. Evaluation . . . 618
   4.3. Analysis and discussion . . . 619
5. Conclusion . . . 619
Acknowledgements . . . 621
Reference . . . 621

1. Introduction

As deterioration of environment, exhaustion of fossil-fuel and increasing demand for electricity, wind power has been attracting significant attention in many countries worldwide. In fact, wind power is the fastest growing source of renewable energy [1]. However, variable nature of wind energy will possibly put the reliability, stability and power quality of the electricity power system at risk [2].

One of the most essential measures to mitigate serious influence from integration of wind farms is the short-term wind power prediction. Accurate and reliable wind power prediction

* Corresponding author. Mob.: +86 159 116 8089.
E-mail addresses: yanjie_freda@163.com, jiejie300@sina.com (J. Yan).

1364-0321/$ - see front matter
http://dx.doi.org/10.1016/j.rser.2013.07.026

Nomenclature

ANN	Artificial neural network
SVM	Support vector machine
RVM	Relevance vector machine
ORVM	Optimized relevance vector machine
NWP	Numerical weather prediction
GA	Genetic algorithm
PSO	Particle swarm optimization
GA–ANN	ANN optimized by genetic algorithm
SOFM	Self-organizing feature mapping
WT	Wind turbine

allows: (1) reasonable maintenance schedules to be established and minimum spinning reserve capacities to be determined so as to reduce operating costs; (2) proportion of wind power in the electric system to be increased; (3) competitiveness of wind power companies to be improved in competitive bidding markets [3,4].

There are many commonly used wind power prediction methods including the physical methods [43,44] (analytical method and CFD method) and the statistical methods (such as artificial neural network [5–8], fuzzy logic [34,40], the Cao algorithm [38], support vector machine [9–11] and ensemble method [39]). And the above prediction models have been applied for facilitating the economical maintenance schedules, competitive bidding markets [3,4] and unit commitment [41,42]. However, they suffer some disadvantages as well. As for the physic method, the analytical method is hard to meet the precision requirement, and the key problem for CFD method is the computational burden. Among the statistical methods, the most widely used ones are ANN in terms of its generalization ability of prediction. Since ANN can theoretically approximate any nonlinear continuous function, it has been successfully applied to the wind power prediction. However, the performance of the ANN based model is sensitive to the size of training samples [37] and only minimizing the training error of a neural network may lead to over-fitting problem [12]. The consequence is that for the known inputs prediction error is very small while for the unknown inputs out of samples the prediction error surges. And this is termed as limited generalization [13]. To remedy this problem, to increase the number of training samples is one way of improving the generalization performance of ANN. However, this demands for large amount of training samples which in turn limits the application of ANN model. For instance, it is difficult for newly-built wind farms having insufficient historical data to build prediction model because of their short running time. Furthermore, the shortage of historical data will undoubtedly increase the difficulty of training prediction model according to weather variation of different months or different seasons.

For the purpose of enhancing generalization ability without size requirement of training samples, a statistical learning technology SVM has been applied in wind power prediction. The SVM employs a linear function in high-dimensional feature space as hypothesis space and makes good predictions using small training samples. This is a highly effective mechanism for avoiding over-fitting. However, despite its success, we can identify some significant and practical disadvantages of SVM [14]:

(1) The kernel function must satisfy Mercer′s condition. That is, it must be the continuous symmetric kernel of a positive integral operator;
(2) Only a single point estimate can be achieved without any uncertainty information;
(3) Although relatively sparse, the number of support vectors grows linearly along with the increase of training samples size, which increases the computational complexity;
(4) It is necessary to estimate some insensitive parameters which generally entail extra calculation and setting of parameters.

To overcome above drawbacks, a probabilistic learning framework termed relevance vector machine (RVM) has been originally introduced by Tipping [14]. RVM is a nonlinear pattern recognition model with simple structure based on Bayesian Theory and Marginal Likelihood. The key feature of RVM is that as well as offering excellent performance of prediction and generalization, it improves the inadequacy of SVM [15–17]. Therefore, this approach has been successfully applied in many fields, such as: load forecasting, fault classification [18–21], but has not yet been applied to wind power prediction.

In addition to advanced mathematics, different strategies have been developed to improve the forecasts accuracy. Many researchers proved the existence of smoothing effect which means that the overall wind power fluctuations would decrease because of the offset of different wind resources in a large area [24,25]. This effect would grow with the increase of wind farm area and could be employed to manage the electricity quality and fluctuations of wind farm output [26–30]. In fact, this effect would happen when forecasts the power output of a single wind farm, because the forecast errors of each wind turbine which locates at different sites could offset with each other and then reduce the whole forecast error of wind farm [31,32]. However, the computational costs would surge if the output of each wind turbine is predicted, especially when NWP of each wind turbine site is needed. Therefore, a grouping method has been developed to divide wind turbines into several groups considering the factors of wind speed correlation, wind power correlation, wind turbine sites. The forecasting model is established for each group and the forecasts of wind farm output is derived by adding results of each group together.

Considering smoothing effects, features of numerical weather prediction (NWP) error and historical data of wind farms and nature of RVM mentioned above, it is not suitable to apply a raw RVM model directly in wind power prediction. Therefore, an optimized relevance vector machine based method (ORVM), a combination of RVM, grouping method and selection method for training samples and particle swarm optimization (PSO) are used to predict wind farm output of each month.

In this paper, Section 2 describes the theory of RVM. Section 3 describes ORVM wind power forecasting model. To verify the effectiveness and superiority of the ORVM model, Section 4 presents a case study with two wind farms comparing the performance of ORVM, SVM and GA–ANN. Section 5 includes the final conclusions.

2. Theory of relevance vector machine [14]

Given a set of input-target pairs $\{x_n, t_n\}_{n=1}^{N}$, assume that $t_i = y(x_i; w) + \varepsilon_i$. ε_i is assumed to be mean-zero Gaussian with variance σ^2, that is $N(0, \sigma^2)$. Kernel function $K(x, x_i)$ has been considered in RVM which makes prediction by the function:

$$y(x;\omega) = w^T \varnothing(x) = \sum_{i=1}^{M} w_i K(x,\ x_i) + w_0 \tag{1}$$

where, $\varnothing(x)$ is vector of basis function; $w = (w_1, w_2, \ldots, w_M)$ is weights vector.

Therefore, the probabilistic formulation of RVM Model is defined as

$$p(t_n|x) = N(t_n|y(x_n), \sigma^2) \quad (2)$$

where, N represents a Gaussian distribution over t_n with mean of $y(x_n)$ and variance σ^2. And the definition of $y(x_n)$ is the same as function (1).

The likelihood function of whole samples is defined as follow:

$$p(t|w, \sigma^2) = (2\pi\sigma^2)^{-N/2} e^{\{-(1/2\sigma^2)||t-\varnothing w||^2\}} \quad (3)$$

To overcome over-fitting from implement of maximum-likelihood estimation for w and σ^2, constraint on weights w_i was imposed, that is ‘prior’ probability distribution as follow:

$$p(w|\alpha) = \prod_{i=0}^{N} N(w_i|0, \alpha_i^{-1}) \quad (4)$$

where, α is N+1 vector termed ‘hyperparameters’.

The posterior over unknown samples could be obtained from proceeds of Bayes inference.

$$p(w, \alpha, \sigma^2|t) = \frac{p(t|w, \alpha, \sigma^2) \times p(w, \alpha, \sigma^2)}{p(t)} \quad (5)$$

Assuming that new test target is t_*, new test input x_* are used to make predictions. Then, predictive distribution can be written as:

$$p(t_*|t) = \int p(t_*|w, \alpha, \sigma^2)\, p(w, \alpha, \sigma^2|t) \mathrm{d}w \mathrm{d}\alpha \mathrm{d}\sigma^2 \quad (6)$$

Posterior distribution over weighs could be consequently rewritten as:

$$p(w|t, \alpha, \sigma^2) = \frac{p(t|w, \sigma^2) \times p(w|\alpha)}{p(t|\alpha, \sigma^2)} \quad (7)$$

Therefore, learning process of RVM becomes the search for α and σ^2 which makes the maximization of $p(\alpha, \sigma^2|t) \propto p(t|\alpha, \sigma^2)\, p(\alpha)p(\sigma^2)$ by maximum marginal likelihood estimation methods

$$p(t|\alpha, \sigma^2) = \int p(t|w, \sigma^2)p(w|\alpha)\mathrm{d}w$$
$$= (2\pi)^{-(N/2)}\ |\sigma^2 I + \varnothing A^{-1}\varnothing^T|^{-(1/2)} \exp\left\{-\frac{1}{2}t^T(\sigma^2 I + \varnothing A^{-1}\varnothing^T)^{-1}t\right\} \quad (8)$$

It can compute α, σ^2 by equating the differentiation over α, σ^2 of function (8) to zero. When α_i approaches extremely large, w_i goes to zero because of constrain by the prior. For w_i interrelated with small α_i, it fits sample data better. Iteration would be preceded until the convergence condition is fulfilled. During the process of parameters estimation, most of $\alpha_i \to \infty$, where corresponding $w_i = 0$. It leads to non-participation of prediction calculation for many terms of kernel matrix. This is why RVM could achieve sparsity.

Iterative estimation of hyperparameters proceeds to make predictions based on each weigh of posterior distribution which adjusts to maximizing values α_{MP}, σ^2_{MP}. With new inputs x_*, predictive results could be described as follow:

$$p(t_*|t, \alpha_{MP},\ \sigma^2_{MP}) = \int p(t_*|w,\ \sigma^2_{MP})\ p(w|t, \alpha_{MP},\ \sigma^2_{MP})dw = N(t_*|y_*,\ \sigma^2_*) \quad (9)$$

where,

$$y_* = \mu^T \varnothing(x_*) \quad (10)$$

$$\sigma^2_* = \sigma^2_{MP} + \varnothing(x_*)^T\ \Sigma\ \varnothing(x_*) \quad (11)$$

3. RVM-based wind power interval forecasting model

3.1. Model structure

Structure of the ORVM wind power grouping forecast model is illustrated in Fig. 1, composed of a grouping engine using SOFM method, a selection engine for training samples and a PSO engine for optimizing parameters of kernel function and a RVM engine for forecasting and its confidential interval.

3.2. Grouping of wind turbines

Wind power production is directly correlated to the wind speed through a power curve. Other atmospheric factors, such as wind direction, pressure, temperature, relative humidity, also have impact on the actual power output [22,23]. In this paper, single NWP results are adopted as inputs of ORVM prediction model, and the reference site of this single NWP is at met mast. Although NWP has already well applied in wind power forecasting, there are also some disadvantages such as representativeness of single reference site, huge computation, low computational resolution etc [33–35]. Current practice is to calculate NWP at met mast to represent the wind profile of the whole wind farm which impacts the accuracy of wind power forecasting. Especially with the increase of wind farm size, the representativeness of a single met mast is

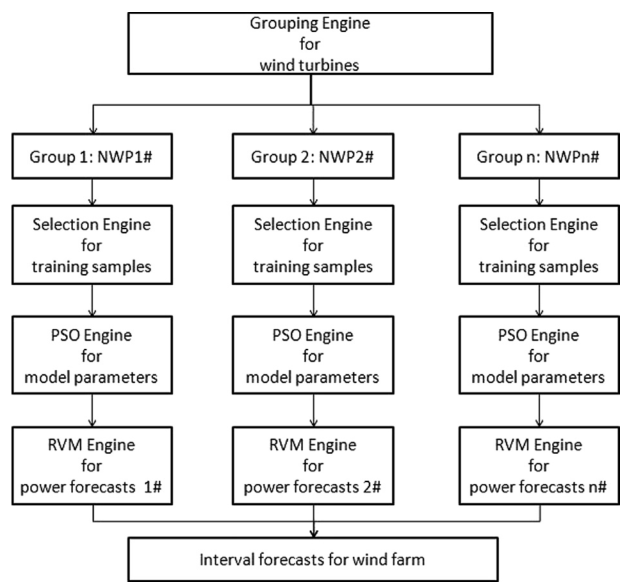


Fig. 1. Structure of the ORVM grouping forecast model.

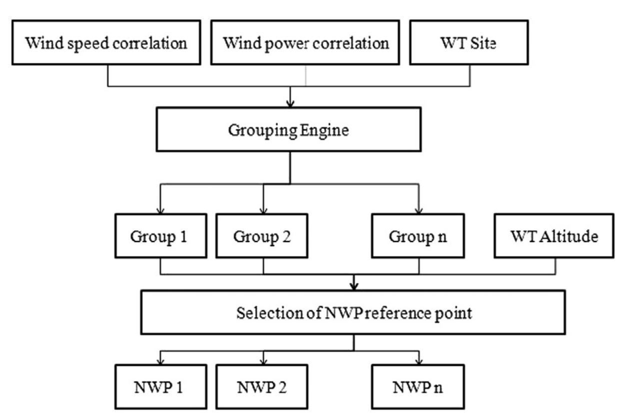


Fig. 2. Structure of the grouping engine.

accordingly weakening. The more the reference sites are selected to predict the weather, the higher the forecasting accuracy might be achieved but the larger computational burden would sequentially bring in. To balance the practical application and forecasting accuracy, a grouping model has been proposed based on a clustering method—SOFM [36] to identify similarity of wind speed, wind turbine generating characteristic, wind turbine site. As shown in Fig. 1, grouping engine would divide wind turbines into n groups and every group would have its own reference site for NWP. Based on NWP of each group, the following engines would subsequently continue to conduct. Note that, the wind turbine which suffers less influence from wake effect, topography in the same group would be selected as reference site of NWP. For example, the wind turbine locates at higher altitude or locates relatively far from other wind turbines.

The structure of grouping engine is shown in Fig. 2. Wind speed correlations, wind power correlation of each wind turbine help the engine put the correlated wind turbines into the same group. Meanwhile site of each wind turbine brings the geographic similarity into the grouping engine. After grouping, the representativeness of wind turbine in its own group would be evaluated by altitude and distance to wind farm border. To take a wind farm in north China (named as WF1) as example, the results of grouping engine and NWP reference point for each group are shown in Table 1.

3.3. Selection of training samples

Figs. 3 and 4 present accuracy of wind speed from NWP whose the reference site is at met mast in two wind farms (WF1 and WF2). Observe that correlation coefficient and root mean square error of NWP wind speed forecasts are not stable and fluctuate with the changes of months or seasons. In Fig. 3, forecasts accuracy of NWP normally peaks in winter (Dec.–Mar.), drops to lowest in spring (Apr.–June), following a modest rise in summer and fall (July–Nov.), but even not reaching the maximum in winter. In Fig. 4, the NWP accuracy shows very similar features: the accuracy peaks in winter and summer and drops to its lowest in autumn and spring. That is because that there are some months or seasons whose weather changes are relatively stable and display strong regularity. For those meteorological phenomena fitting into our known regularity, its NWP could be more accurate, otherwise it is hard to predict precisely for complex weather variation.

Table 1
Results of grouping engine.

Group	Reference site of NWP	WT in the group
1	WT 58#	16,17,20,55,56,58
2	WT 18#	15,18,69,70,76
3	WT 21#	14,19,21,57
4	WT 24#	23–27,32–36,44,45,54
5	WT 63#	63
6	WT 67#	62,64,65,67,68,74,75
7	WT 59#	22,59,60,82
8	WT 116#	40,116,119
9	WT 43#	28–31,37–39,41–43,46–49,53,112,114
10	WT 12#	12
11	WT 73#	71,73
12	WT 87#	61,80,85,87
13	WT 102#	101,102
14	WT 52#	51–52,103,107,109–111,113,115,117–122
15	WT 72#	66,72,77–79,81,83,84,86
16	WT 99#	88–100
17	WT 108#	9,108
18	WT 8#	1–8,10,11,13

Note: Reference site of NWP indicates the location on which the NWP data are calculated.

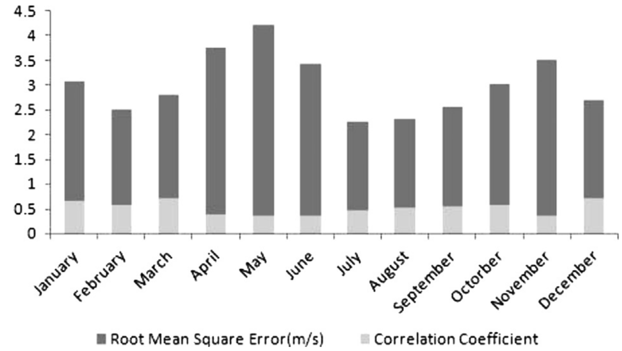


Fig. 3. Accuracy of wind speed from NWP for each month in WF1.

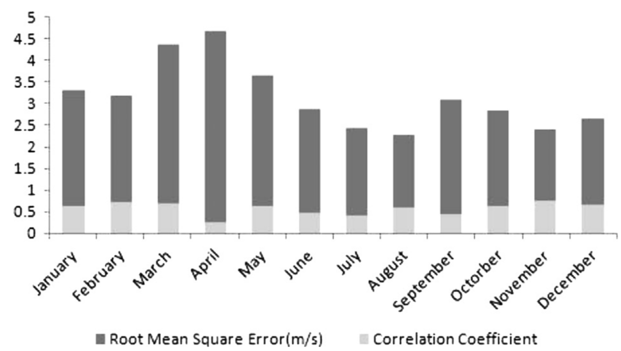


Fig. 4. Accuracy of wind speed from NWP for each month in WF2.

As weather forecast within wind farms is the key of wind power prediction technology and the NWP accuracy shows clear monthly or seasonal characteristic, it is possible to improve the accuracy of wind power forecasting if models of each month are built. Furthermore, RVM has the advantages of demand of less training samples, which facilitates the model training for each month. Thus, candidate NWP samples of each month are transferred to the selection engine as model inputs and then to build prediction models according to different error distribution of NWP.

Since RVM is an intelligent learning method with adaption of small training samples, it is necessary and feasible to select the most beneficial training data for accurate prediction. On the one hand, selection of training samples exerts a significant impact on forecast accuracy. A small number of training samples or samples with large deviation usually causes that forecast engine learns mapping function incorrectly while a huge amount of training samples may be misleading for the forecast engine. On the other hand, data in a month from wind farms are far more than the required amount of training samples for RVM, because data resolution of wind farms is normally no more than 15 min, and there are theoretically around 2880 sets of data per month. Therefore, how to select the most effective NWP data sets as the training samples is the key step in predictive modeling.

To test WF1 (January 2010) and WF2 (January 2011), the accuracy of NWP is classified according to absolute error of NWP wind speed. For instance, <1.2 m/s means higher NWP accuracy than <1.8 m/s because only samples whose absolute error of NWP wind speed is less than 1.2 m/s rather than 1.8 m/s are selected as training samples. It starts with a low accuracy level of NWP and then gradually increases it. As shown in Fig. 5, two lines show similar trend meaning that the accuracy of forecasting model increases with the rise of NWP accuracy at first. That is because that with improvement of NWP forecasts, training samples might simulate actual power curve of wind turbines more accurately so as to achieve higher wind power forecasting accuracy. In this case

of WF1, forecasting accuracy peaks at the point of absolute error of NWP wind speed <1.5 m/s, while it is <1.6 m/s in WF2. Then, forecasting accuracy declines from the highest value with further descent of NWP Error. Because with further increase of NWP accuracy, elimination of too much data leads to too small training samples so that deteriorate both the generalization and fitness of forecasting model as well as forecast accuracy.

This selection engine was implemented with candidate samples of each month in two wind farms and the results were recorded in Table 2. Forecast accuracy of each month reflects similar features as that of January. Besides, accuracy peak of forecasting model for different months locates at different absolute error of NWP wind speed (or NWP accuracy level). Considering using samples corresponding to different accuracy level of NWP results in different accuracy of forecasting model, selection of training samples according to NWP accuracy level may improve the forecast accuracy.

3.4. Optimization of kernel function

Due to the significant impact of kernel function parameters on forecasting accuracy, particle swarm optimization (PSO) has been adopted to search for the optimal kernel width and initial value of RVM. In this paper, Gaussian kernel function as (12) is adopted.

$$K(x,x_i) = \exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right) \tag{12}$$

where, σ is the width of kernel function.

There are 20 particles for each parameter used in this model and the adaptive function of them is root mean square error. The speed and location of them are updated by following functions:

$$v_{i,d}^{k+1} = \omega v_{i,d}^{k} + c_1 rand()(pb_{i,d}^{k} - x_{i,d}^{k}) + c_2 rand()(gb_{i,d}^{k} - x_{i,d}^{k}) \tag{13}$$

$$x_{i,d}^{k+1} = x_{i,d}^{k} + v_{i,d}^{k+1} \tag{14}$$

where, c_1 and c_2 are learning factors; $rand()$ is uniform random number [0, 1]; $v_{i,d}^{k}$ and $x_{i,d}^{k}$ are speed and location of the ith particles in the kth iteration in d-demention; $pb_{i,d}^{k}$ and $gb_{i,d}^{k}$ are, respectively, the individual best location and group best location of the ith particle in d-demention; ω is inertia weight factor.

Along with the picked training samples of one month and its corresponding actual power data, optimized parameters from PSO engine are transferred to RVM engine whose structure is illustrated in Fig. 6.

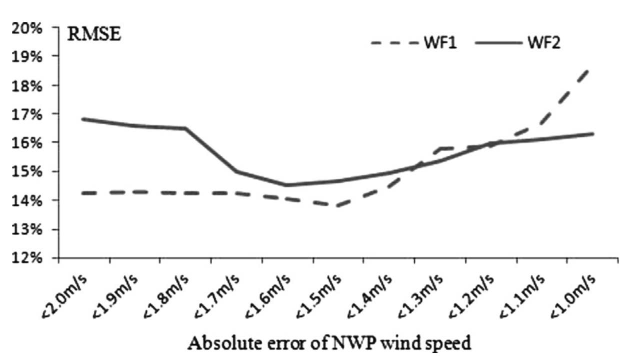


Fig. 5. Forecasts accuracy corresponding to different NWP error level.

3.5. Steps of grouping forecasts

Application of the ORVM model for wind power prediction can be summarized as following steps:

(1) Wind turbine grouping phase:
Assume that M is the size of input samples; N denotes the size of the number of neural cell in output layer; k represents the neural cell number in input layer; t is the current number of learning times; T is the total learning times.

a. To calculate the correlation of wind speed and wind power output of each wind turbine;
b. To input the wind turbine sites and the correlation of wind speed and wind power into grouping engine as input vectors and to normalize the input vectors among the range of [0,1] as $X=\{x_i; x_i=(x_{i1}, x_{i2}, \ldots, x_{ik})^T \in R^k,\ i=1,2,\ldots,M\}$
c. To initialize the SOFM based grouping engine: connection weights $W=\{w_j;\ w_j=(w_{j1},\ w_{j2}, \ldots, w_{ji}, \ldots, w_{jk})^T \in R^k,\ j=1,2,\ldots,N\}$; learning pace $\alpha(t)$; neural cell neighborhood $Na_j(t)$;
d. To calculate the distance between connection weight of neural cell and inputs:

$$Ed_j = \left[\sum_{i=1}^{k} (x_i - w_{ji})^2\right]^{1/2},\ j=1,2,\ldots \tag{15}$$

e. To search the winning neural cell which has the shortest distance:

$$Ed_{j*} = \min_{1 \le j \le N} \{Ed_j\} \tag{16}$$

f. To adjust the network parameters (connection weights of the winning neural cell with other neural cell) in order to mapping the input vectors into the

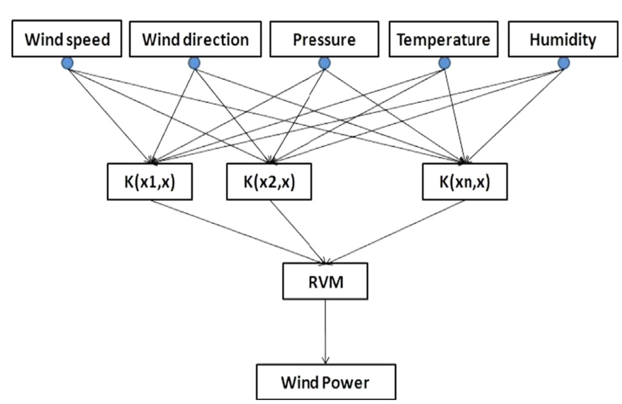


Fig. 6. Structure of raw RVM engine.

Table 2
NWP accuracy level for the training sample selection of each month in two wind farms.

(m/s)	January	February	March	April	May	June	July	August	September	October	November	December
WF1	< 1.5	< 1.0	< 1.5	< 1.5	< 1.6	< 2.0	< 2.0	< 1.6	< 1.2		< 1.7	< 1.5
WF2	< 1.6	< 1.5	< 1.7	< 1.9	< 1.6	< 2.0	< 1.2	< 1.2	< 1.5	< 1.6	< 1.6	< 1.7

output layer:

$$w_j(t+1) = w_j(t) + \alpha(t)[x(t) - w_j(t)],\ j \in Na_{j*}(t) \quad (17)$$

g. To iterate every sample and to update the learning pace until $t = T$:

$$\alpha(t) = \alpha(0)\left(1 - \frac{t}{T}\right) \quad (18)$$

h. Each input vector would non-linearly mapping to their winning neural cell in output layer. The vectors mapping to the same winning neural cell belong to the same category.

(2) Training samples selection phase:
To classify the NWP accuracy level and to input the overall candidate samples of each month into selection engine to determine the most effective level. Samples selected by the most effective level were taken as training samples. Then, to normalize selected samples into range of 0 to 1 for the convenience of computation.

(3) Model parameters optimization phase:
To transfer the selected training samples from phase (1) to PSO engine to determine the most suitable kernel width and initial value of RVM iteration.

(4) Training and forecasting phase:

a. To input selected training samples and optimized parameters into RVM engine and to initialize iteration conditions;

b. To calculate the posterior distribution over weights:

$$p(w|t,\alpha,\sigma^2) = (2\pi)^{-(N+1/2)}|\Sigma|^{-(1/2)}e^{\{-(1/2)(w-\mu)^T\Sigma^{-1}(w-\mu)\}} \quad (19)$$

c. To update the mean and variance of posterior μ and Σ, respectively:

$$\mu = \sigma^{-2}\Sigma\varnothing^T t \quad (20)$$

$$\Sigma = (\sigma^{-2}\varnothing^T\varnothing + A)^{-1} \quad (21)$$

d. To precede iteration of function (22) and (23) until the convergence condition is fulfilled:

$$\alpha_i^{\text{new}} = \frac{1 - \alpha_i\Sigma_{ii}}{\mu_i^2} \quad (22)$$

$$(\sigma^2)^{\text{new}} = \frac{t - \varnothing\mu^2}{N - \Sigma_i(1 - \alpha_i\Sigma_{ii})} \quad (23)$$

where, μ_i is the ith mean of posterior from function (20); Σ_{ii} is the ith diagonal element of posterior covariance from function (21), and computed by α, σ^2 from current iteration results; N indicates the number of sample data.

e. To delete those $w_i = 0$ in the iteration process. The vector corresponding to remaining w_i is termed as 'Relevance Vector'. Consequently, RVM largely reduces model complexity and computing costs. Model parameters α_{MP}, σ^2_{MP} would be achieved when model training is completed.

f. To import test data and model parameters α_{MP}, σ^2_{MP} into forecasting model. After calculation and anti-normalization process, prediction value and its possible fluctuation are obtained.

Not only could ORVM model provide an individual prediction value, but also calculate variance of the prediction value by function (11) which is also the possible prediction error. Term of variance comprises two kinds of error: one is from error of estimation while another is from uncertainty of weights calculation. This probabilistic mechanism for forecasting largely improves practical value for risk-resisting [14].

4. Case study

4.1. Data

The data of two wind farms in China include mean wind farm output collected from SCADA and mean wind speed from met mast and numerical weather prediction data. All the data had an interval period of 15 min covering running period shown in Table 3. In WF1, wind speed data measured by each wind turbine SCADA are available while it is not available in WF2. It means that grouping engine could only be tested in WF1. Among the available data, 80% are considered as candidate training samples and 20% test samples.

To evaluate the performance of proposed model, ORVM models are compared with support vector machine (SVM) and ANN optimized by Genetic Algorithm (GA–ANN) in terms of forecast accuracy, model complexity and model running time. For the sake of fairness, all methods have the same input variables, training samples and test samples. Note that ANN′s training samples contains 12 sets of ORVM training samples as a whole due to its demand of a large number of training samples which means only one model established.

4.2. Evaluation

Two frequently used error criteria are adopted for numerical experiments of this paper. Root mean square error (RMSE) in function (24) is computed for all validation period and can give a better evaluation of prediction error over a longer period [12]. Mean absolute error (MAE) in function (25) is another commonly used error measures for prediction process.

$$\text{RMSE} = \frac{\sqrt{\sum_{i=1}^{n}(P_{Mi} - P_{Pi})^2}}{Cap \times \sqrt{n}} \quad (24)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|P_{Mi} - P_{Pi}|}{Cap \times n} \quad (25)$$

where P_{Mi} and P_{Pi} indicate actual and forecast value of wind power output at time of i; Cap denotes installed capacity of wind farm; n is number of samples involved.

Table 3
Wind farm description.

	WF1	WF2
Installed capacity	183 MW	201 MW
Running period	2010 year except for October	2011.5–2012.6
WT number	122	134
Location	Northeast China	East-central China

Table 4
RMSE comparison of single NWP forecasts and grouping forecasts in WF1.

	ORVM (%)	SVM (%)	GA–ANN (%)
Single NWP	9.9	12.5	13.3
Grouping forecasts	9.1	11.4	12.1

Table 5
Comparisons of forecasts accuracy for each month in WF1 (forecasts with grouping engine).

Month	ORVM model		SVM model		GA–ANN model	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Average	0.091	0.059	0.114	0.082	0.121	0.091
January	0.109	0.077	0.112	0.076	0.127	0.094
February	0.103	0.061	0.126	0.093	0.124	0.099
March	0.077	0.049	0.121	0.096	0.091	0.061
April	0.137	0.100	0.139	0.117	0.133	0.090
May	0.103	0.071	0.175	0.145	0.149	0.115
June	0.063	0.035	0.056	0.035	0.090	0.069
July	0.071	0.044	0.071	0.041	0.112	0.084
August	0.050	0.041	0.064	0.032	0.089	0.06
September	0.086	0.053	0.098	0.056	0.143	0.103
November	0.101	0.054	0.170	0.121	0.151	0.138
December	0.098	0.072	0.123	0.09	0.120	0.092

Table 6
Comparisons of forecasts accuracy for each month in WF2 (forecasts without grouping engine).

Month	ORVM model		SVM model		GA–ANN model	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Average	0.119	0.092	0.144	0.104	0.142	0.106
January	0.140	0.090	0.147	0.092	0.159	0.101
February	0.137	0.091	0.141	0.097	0.150	0.119
March	0.142	0.120	0.149	0.125	0.147	0.125
April	0.151	0.134	0.177	0.1403	0.162	0.127
May	0.169	0.146	0.174	0.157	0.185	0.151
June	0.142	0.111	0.155	0.107	0.169	0.140
July	0.083	0.05	0.141	0.084	0.128	0.079
August	0.098	0.073	0.169	0.108	0.148	0.097
September	0.101	0.081	0.128	0.091	0.114	0.087
October	0.053	0.038	0.076	0.050	0.075	0.051
November	0.101	0.070	0.152	0.111	0.148	0.107
December	0.113	0.103	0.125	0.095	0.127	0.097

4.3. Analysis and discussion

Table 4 shows the grouping forecasts of three models to verify the effectiveness of grouping engine. Single NWP represents the forecasting using only one set of NWP data at met mast while grouping forecasting indicates the forecasting using several sets of NWP data at reference sites of each group in Table 1. The results show that the grouping engine plays positive role on three different forecasting methods, and improve the yearly average RMSE by (9.9 − 9.1%)/9.9% = 8.08% for ORVM, (12.5 − 11.4%)/12.5% = 8.8% for SVM, (13.3 − 12.1%)/13.3% = 9.02% for GA–ANN.

Tables 5 and 6 show the full-year forecasts accuracy of ORVM, SVM and GA–ANN in WF1 and WF2. Because of the data management in WF2, wind speed data measured by wind turbines are unavailable. Consequently, the grouping engine could be tested in WF1 but not in WF2. That is one of the reasons why the performance of ORVM in WF1 is better than that of in WF2. Moreover, a little worse NWP quality in WF2 shown in Figs. 3 and 4 might be another reason for larger forecasting RMSE. In general, RMSE and MAE defined in functions (24) and (25) of ORVM are considerably lower than those of SVM and GA–ANN. Moreover, the average RMSE and average MAE are less than those of SVM by about (11.9 − 9.2%)/11.9% = 22.68% and (8.2 − 5.9%)/8.2% = 28.04% in WF1; (14.4 − 11.9%)/11.4% = 21.92% and (10.4 − 9.2%)/10.4% = 11.53% in WF2; revealing capability of ORVM model in wind power prediction.

To better illustrate the forecast trend and probabilistic forecast capability of the proposed model, forecasting results of four days from different seasons are presented as examples of each wind farm. Figs. 7 and 8 show the predictive and real value of wind farm output as well as the range of possible fluctuations at 90% confidence level. ORVM features in providing not only a certain predictive value but also its uncertainty analysis, both the upper and lower limits of power fluctuation. This can provide more scientific guidance of risk decision for both wind farm operators and electric system dispatchers.

Table 7
Comparisons of computing time and vector number for each model.

	Model	Training time (s)	Test time (s)	Number of vectors involved
WF1	ORVM	192.136	52.190	92.52
	SVM	180.119	67.529	116.39
	GA–ANN	1286.342	109.327	
WF2	ORVM	15.354	0.962	69.25
	SVM	12.008	1.003	93.06
	GA–ANN	513.452	1.795	

Validation test of uncertainty analysis: throughout the year, the percentage of which real power production locates within the fluctuation range is 89.928% on average in WF1, 88.31% in WF2 at 90% confidence level. It proves that ORVM model can quantitatively assess the wind power prediction uncertainty.

To evaluate efficiency of the proposed model, its computing time (containing training and test time) and vector number are presented in Table 7 as well as those of SVM and GA–ANN. Less computational cost and few involved vectors reflect its efficient learning capacity and simple model structure. The running time, measured on a simple hardware set of 2.79 GHz Processor with 3.12 GB RAM, is completely acceptable for decision-making of a day-ahead dispatching or even for ultra-short term operating. Note that the training and running time of WF1 largely outnumber those of WF2 because the test in WF1 involves grouping engine meaning that more sets of NWP data and more forecasting models join in the calculation process.

5. Conclusion

1. ORVM models of each month for wind power grouping prediction have been proposed. Results of the case study involving two wind farms in China prove that full-year average RMSE and MAE of the proposed model are 9.1% and 5.9% which are lower than those of SVM and GA–ANN, respectively. Furthermore, the proposed model effectively provides quantitative assessment of forecasting uncertainty. ORVM model outperforms SVM and GA–ANN in terms of the wind power forecast accuracy and practicality.
2. A grouping engine has been established to divide wind turbines into several groups for improving forecasting accuracy and to minimize NWP computational cost with the smoothing effect. The grouping engine can identify the similarity of distinctive wind speeds, WT power outputs and wind turbine sites at different locations. And then a NWP reference site of each group will be selected to represent the general condition of the wind resource of each group. This will help enhance the representativeness of NWP data in a certain area so as to improve forecasting performance.
3. A method for selection of training samples is presented considering instability and seasonal characteristic of NWP accuracy as well as requirement of small training samples of RVM. This method makes forecasting models more suitable to NWP characteristics of each month so as to significantly improve forecast accuracy. Besides, PSO has been applied to search the appropriate model parameters for different samples.

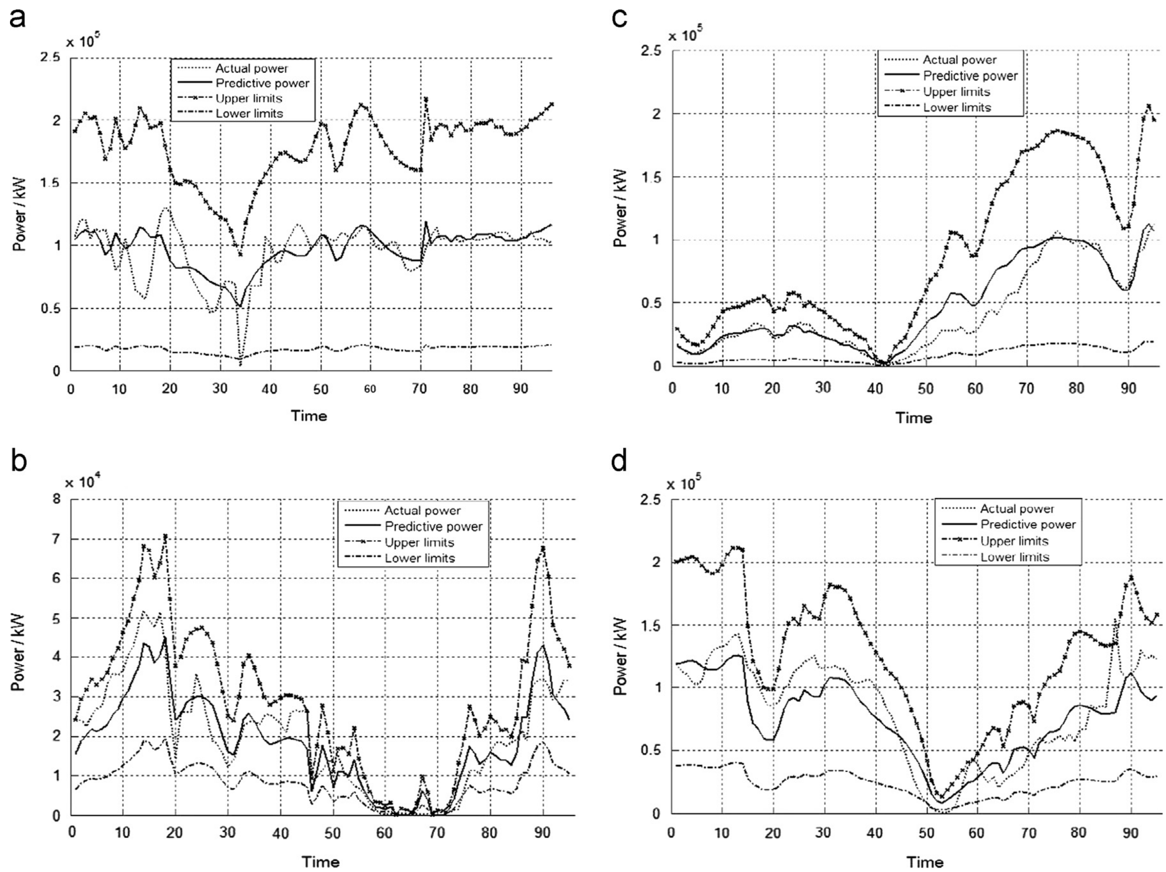


Fig. 7. Probabilistic forecast results of WF1 in 24th May (a), 29th July (b), 25th Sep. (c), 26th Dec. (d).

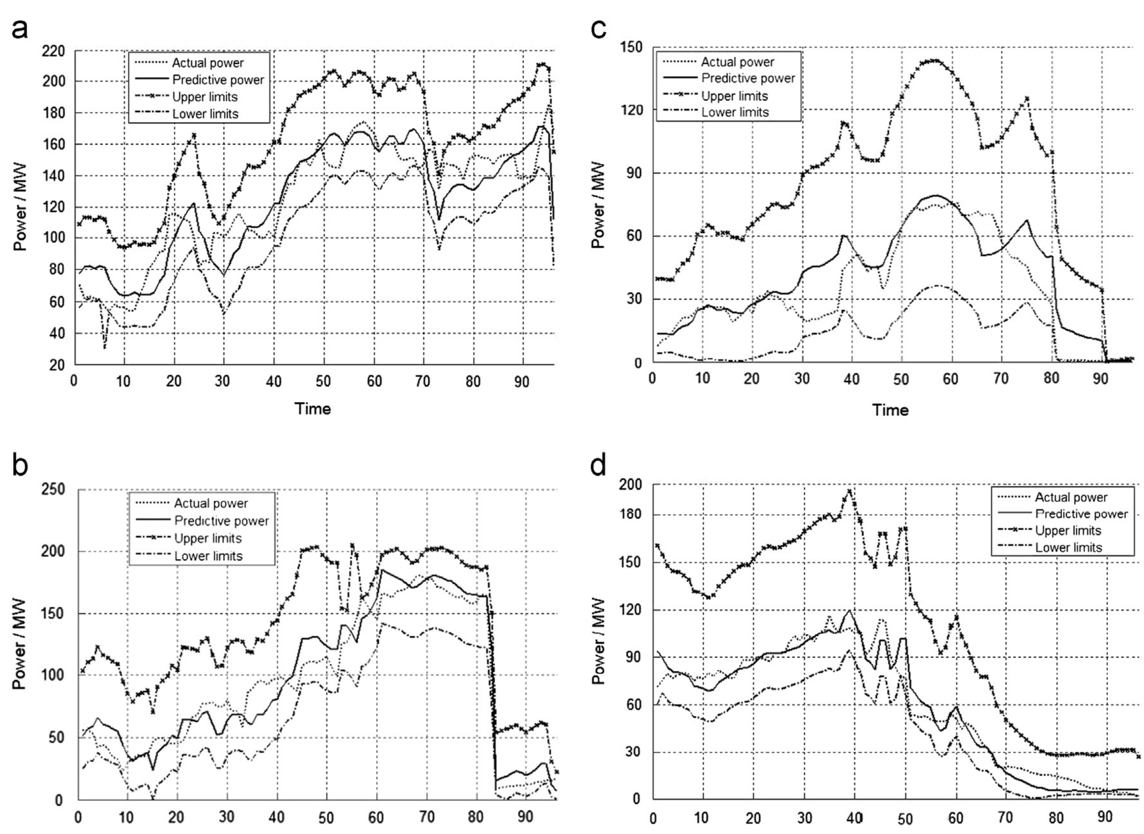


Fig. 8. Probabilistic forecast results of WF2 in 15th Feb. (a), 20th April (b), 15th July (c) and 17th Dec. (d).

4. The merits of RVM-based model for wind power prediction are as follows:
   - Diverse selection of kernel function improves model adaptability: There is no necessity to satisfy Mercer's condition for kernel function. Thus, RVM-based model could more precisely simulate power output of different wind farms in a wider scope;
   - Probabilistic prediction: RVM-based model provides fluctuation range of prediction at given confidence level rather than a certain value of wind power prediction;
   - Fewer samples required in training process: On the one hand, due to facilitation of building prediction models of each month, it reflects characteristic of NWP accuracy distribution more properly and then enhances prediction capacity. On the other hand, it is capable of building prediction model for newly-built wind farms which have less historical data;
   - Sparsity: Most relevance vectors automatically tend to zero during training process. Consequently, it has much less vectors than that of SVM in the computation. Moreover, number of relevance vectors would not suffer linear increase along with growth of size of training samples. Because of the focus on vectors only relating with accurate prediction, model complexity is greatly reduced and training efficiency is improved;
   - Simplified parameters setting: Different from SVM, only the width of kernel function being set minimizes subjective influence upon RVM-based model to the largest extent.

Acknowledgements

This work is supported by project granted by Chinese National Energy Administration (NY20110204-1) and Chinese State 863 key grant project for Smart Grid (SQ2010AA0523193001).

Reference

[1] Taylor JW, McSharry PE, Yokyama R, Chen L, Lee W-J. Forecasting the wind generation using a two-stage network based on meteorological information. IEEE Transactions on Energy Conversion 2009;24(2):474–82.
[2] Usaola J. Probabilistic load flow with correlated wind power injections. Electric Power Systems Research 2010;80(5):528–36.
[3] Lerner J, Grundmeyer M, Garvert. M. The importance of wind forecasting. Renewable Energy Focus 2009;10210(2):64–6.
[4] Hong YY, Chang HL, Chiu. CS. Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs. Energy 2010;35:3870–6.
[5] Mabel MC, Fernandez E. Analysis of wind power generation and prediction using ANN: a case study. Renewable Energy 2008;33(5):986–92.
[6] Cadenas E, Rivera W. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. Renewable Energy 2009;34(1):274–8.
[7] Ramirez-Rosado IJ, Fernandez-Jimenez LA, Monteiro C, Sousa J, Bessa R. Comparison of two new short-term wind-power forecasting systems. Renewable Energy 2009;34(7):1848–54.
[8] Catalão JPS, Pousinho HMI, Mendes. VMF. Short-term wind power forecasting in Portugal by neural networks and wavelet transform. Renewable Energy 2011;36:1245–51.
[9] Liu Yongqian, Shi Jie, Yang Yongping, Han Shuang. Piecewise support vector machine model for short-term wind-power prediction. International Journal of Green Energy 2009;6:479–89.
[10] Mohandes MA, Halawani TO, Rehman S, Hussain Ahmed A. Support vector machines for wind speed prediction. Renewable Energy 2004;29:939–47.
[11] Bouzgou Hassen, Benoudjit Nabil. Multiple architecture system for wind speed prediction. Applied Energy 2011;88:2463–71.
[12] Amjady Nima, Keynia Farshid, Zareipour Hamidreza. Short-term wind power forecasting using ridgelet neural network. Electric Power Systems Research Dec. 2011;81(12):2099–107.
[13] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power. Renewable and Sustainable Energy Reviews 2009;13(4):915–20.
[14] Tipping. ME. Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research 2001;1:211–44.
[15] WM Yu, TH Du, K Lim. Comparison of the support vector machine and relevant vector machine in regression and classification problems [C]. In: Eighth international conference on control, automation, robotics and vision, vol. 2. Kunming: China; Dec. 6–9, 2004.p. 1309–1314.
[16] Bowd C, Medeiros FA, Zhang ZB, et al. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements [J]. Investigative Ophthalmology & Visual Science 2005;46(4):1322–9.
[17] Widodo A, Kim EY, Son JD, Yang BS, Tan ACC, Gu DS, et al. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. Expert Systems with Applications 2009;36:7252–61.
[18] Duan Q, Zhao JG, Niu L, et al. Regression based on sparse Bayesian learning and the applications in electric systems [c]. In: Fourth international conference on natural computing. Jinan: China; Oct. 18–20, 2008. p. 106–111.
[19] Duan Q, Zhao JG, Ma Y. Relevance vector machine based on particle swarm optimization of compounding kernels in electricity load forecasting. Electric Machines and Control 2010;14(6):34–8.
[20] Caesarendra Wahyu, Widodo Achmad, Yang Bo-Suk. Application of relevance vector machine and logistic regression for machine degradation assessment. Mechanical Systems and Signal Processing 2010;24:1161–71.
[21] Lou JG, Jiang JH, Shuai CY, Zhang R, Jin A. Software reliability prediction model based on relevance vector machine. Intelligent Computing and Intelligent Systems, ICIS 2009 2009;20–22:229–33 (Nov.).
[22] Lazic' Lazar, Pejanovic' Goran, Živkovic´ Momčilo. Wind forecasts for wind power generation using the eta model. Renewable Energy 2010;35:1236–43.
[23] Grazia Maria, Giorgi De, Ficarella Antonio, Tarantino Marco. Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. Energy 2011;36:3958–78.
[24] Wei LIU. Research on optimal dispatching of the wind farm cluster [D]. Beijing: Beijing Jiaotong University; 2011.
[25] Sørensen P, Cutululis NA. Wind farms' spatial distribution effect on power system reserves requirements. In: ISIE 2010-IEEE international symposium on industrial electronics. Bari, Italy: IEEE; 2010. p. 2505–2510.
[26] Shimamura T , Yamashita D, Koyanagi K, Nakanishi Y, Yokoyama R. Evaluation of smoothing effect of wind power generator aggregation on power system operation. In: International conference on renewable energies and power quality (ICREPQ'13) Bilbao (Spain); 20th–22nd March, 2013.
[27] Asari M, Nanahara T, Maejima T, Yamaguchi K, Sato. T. A study on smoothing effect on output fluctuation of distributed wind power generation. IEEE Power & Energy Society 2002:938–43.
[28] Li Pei, Banakar Hadi, Keung Ping-Kwan, Golestani Far Hamed, Ooi Boon-Teck. Macromodel of spatial smoothing in wind farms. IEEE Transactions on Energy Conversion 2007;22(1):119–28 (MARCH).
[29] Tarroja B, Mueller F, Eichman JD. Spatial and temporal analysis of electric wind generation intermittency and dynamics. Renewable Energy 2011;36(12):3424–32.
[30] De HJES, Frunt J, Kling KL Mitigation of wind power fluctuations in smart grids. In: Innovation smart grid technologies conference Europe (ISGT Europe). Gothenburg, Sweden: IEEE; 2010.p. 1–8.
[31] Zhi Li, Xueshan Han, Li Han. An ultra-short-term wind power forecasting method in regional grids. Automation of Electric Power System 2010;34 (7):90–4 (in Chinese).
[32] Han Yu, Chang Liuchen. A study of the reduction of the regional aggregated wind power forecast error by spatial smoothing effects in the Maritime Canada. In: IEEE electrical power & energy conference. Halifax, NS: Canada; 2010. p. 942–947.
[33] Efe, B, Unal, E, Mentes, S, Ozdemir, T. Turkey by using the WRF model coupled to WindSim. In: Renewable energy research and applications (ICRERA), 2012 international conference on 11–14 Nov; 2012.
[34] Yang Zhiling, Liu Yongqian, Li Chengrong. Interpolation of missing wind data based on ANFIS, Renewable Energy, 36, 3, 993–998.
[35] Damousis I G, Alexiadis M C, Theocharis J B, Dokopoulos P S. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. IEEE Transactions on Energy Conversion 2004;19(2):352–61 (Jun.).
[36] Kohonen. T. The self-organizing map. Proceedings of the IEEE 1990;78:1464–80.
[37] Jie YAN. Research on uncertainty analysis method for wind power prediction and its application. Beijing; 2012.
[38] Zhang Jin-Hua, Liu Yong-Qian, Tian De. The short-term wind power forecast based on phase-space reconstruction and neural networks. Journal of Digital Information Management 2013;11(1):40–45.
[39] Han Shuang, Liu Yongqian, Yan Jie. Neural network ensemble method study for wind power prediction. In: Proceedings of 2011 Asia-Pacific power and energy engineering conference. Wuhan: China, March 25–28; 2011.
[40] Pinson P, Kariniotakis G N. Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. IEEE Bologna PowerTech - Conference Proceedings 2003;2.
[41] Fabbri A, Roman T Gomez San, Rivier Abbad J, Quezada V H M. Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. IEEE Transactions on Power Systems 2005;20(3):1440–6.
[42] Methaprayoon K, Yingvivatanapong C, Lee W J, Liao J R. An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty. IEEE Transactions on Industry Applications 2007;43(6):1441–8 (November–December).
[43] Li LI, Yongqian LIU, Yongping YANG, Shuang HAN. Short-term wind speed forecasting based on CFD Pre-calculated flow fields. Proceedings of the CSEE 2013;33(7):27–32.
[44] Landberg. L. A mathematical look at a physical power prediction model. Wind Energy 1998;1(1):23–8.